

AD-A130 177

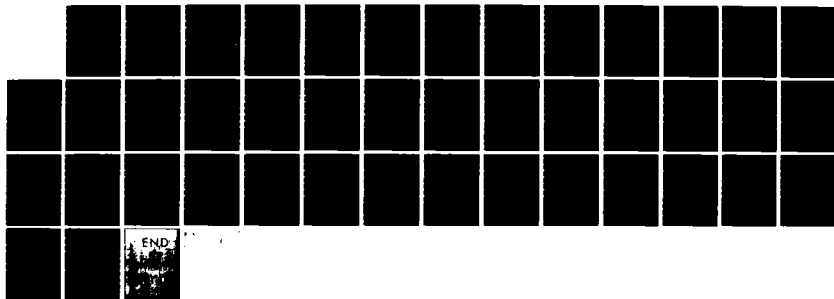
JURY SIZE AND JURY VERDICTS(U) STANFORD UNIV CA DEPT OF  
STATISTICS H SOLOMON 28 JUN 83 TR-335 N00014-76-C-0475

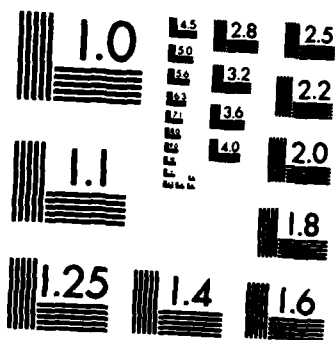
1/1

UNCLASSIFIED

F/G 5/4

NL





MICROCOPY RESOLUTION TEST CHART  
NATIONAL BUREAU OF STANDARDS-1963-A

ADA130177

JURY SIZE AND VERDICTS

BY

HERBERT SOLOMON

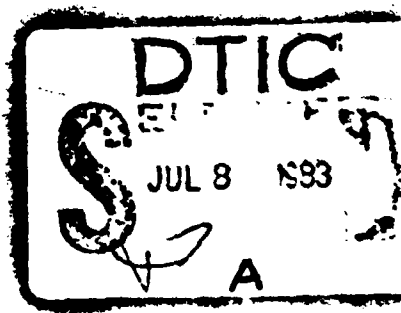
TECHNICAL REPORT NO. 305

JUNE 28, 1933

PREPARED UNDER CONTRACT  
N00014-76-C-0475 (NR-042-267)  
FOR THE OFFICE OF NAVAL RESEARCH

Reproduction in Whole or in Part is Permitted  
for any purpose of the United States Government.  
Approved for public release; distribution unlimited.

DEPARTMENT OF STATISTICS  
STANFORD UNIVERSITY  
STANFORD, CALIFORNIA



83 07 07 051

FILE COPY

JURY SIZE AND JURY VERDICTS

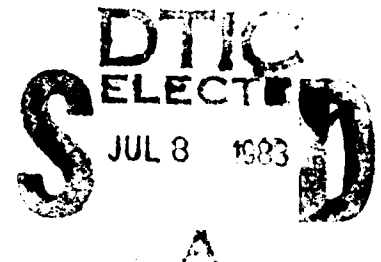
by

HERBERT SOLOMON

TECHNICAL REPORT NO. 335

JUNE 28, 1983

Prepared Under Contract  
N00014-76-C-0475 (NR-042-267)  
For the Office of Naval Research



Reproduction in Whole or in Part is Permitted  
for any purpose of the United States Government.  
Approved for public release; distribution unlimited.

DEPARTMENT OF STATISTICS  
Stanford University  
Stanford, California



Accession For	
NTIS GRA&I	<input checked="checked" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By	
Distribution/	
Availability Codes	
Avail and/or	
Dist	Special
A	

## JURY SIZE AND JURY DECISION\*

Herbert Solomon

### ABSTRACT

↙ The analysis of jury size and jury verdicts in criminal matters now has a long, though interrupted history. Work on this subject in the 18th and 19th centuries by Condorcet and Laplace is discussed and the Poisson model of the 1830's is highlighted. The latter is modified to analyze the American jury experience of the 20th century. Recent U.S. Supreme Court decisions in the 1970's on jury size and jury decision-making have created a resurgence of interest especially on a comparison of six member and twelve member juries. Some comparisons of size in terms of probabilities of errors in verdicts are presented.

↖  
\*Based on an Invited Talk sponsored by the American Statistical Association, the Office of Naval Research, the French Embassy and George Washington University given in Washington, D.C. at a Conference celebrating the 200th anniversary of the birth of Simeon D. Poisson.

### INTRODUCTION

Studies looking into the association of jury size and jury verdicts appear frequently these days. Articles analyzing some

aspects of this subject can be found in journals and books produced for legal scholars, social psychologists, political scientists, statisticians and others. Several articles are essentially reviews or surveys of existing literature even though the time span for published articles in this field is relatively short, see for example Penrod and Hastie (1979). In a report to the Federal Judicial Center, Saks (1981) reviews and analyzes a large literature on small group research and the application of the results to American jury behavior. Sometimes, one jury size model is subjected to an extensive critique in terms of its relevance and approximation to reality as in Kaye (1980).

A series of U. S. Supreme Court decisions on jury size and jury decision rules in criminal cases beginning in 1970 with Williams v. Florida were the motivation for this burgeoning industry. Other decisions soon followed; in 1972 (Johnson v. Louisiana, Apodaca v. Oregon), in 1978 (Ballew v. Georgia) and in 1979 (Burch v. Louisiana). Williams permitted six jurors in state felony trials (reserving twelve for federal felony trials); the next two decisions permitted jury verdicts based on nine out of twelve and ten out of twelve majorities in state felony trials, but then Ballew ruled unconstitutional a jury of size five in a state felony trial; and Burch ruled unconstitutional a five out of six majority decision in a state felony trial.

Thus, in the decade spanning the 1970's important decisions were rendered on jury size and majority requirements for decisions. For the first time, in a very public way, twelve-member juries and unanimity were no longer sacrosanct in felony trials. Smaller juries and majority verdicts had been in existence a long time but they had not been challenged. Some social science writers have referred to the evolving Supreme Court position on jury size as being on a 'slippery slope'. More studies and probably more Supreme Court decisions will follow. The closest precursor to this kind of scholarly, legislative, and judicial activity took place in France at the turn of the 19th century and continued for almost

50 years. The French school of mathematicians engaged in probability theory examined jury size and the jury as a decision making body both from a theoretical and an empirical point of view. Among them were eminent savants such as Condorcet, Laplace, Poisson and Cournot.

Today there is an extensive literature published in the last decade reporting on empirical studies of jury size and jury decision making. A prolific contributor along these lines is Davis and his colleagues and students at the University of Illinois (1973, 1975, 1977, 1978). In this article we will touch on these studies but emphasis will be given to probabilistic models of jury size and jury behavior. The number of investigators in this subject is much smaller than those engaged in empirical efforts.

Those of us who engage in probabilistic models owe a debt to S. D. Poisson and his pioneering work on this topic published in his 1837 book "Recherches sur la probabilité des jugements en matière criminelle et en matière civile". This work contains a detailed and somewhat discursive exposition of a jury behavior model motivated and supported by data on jury trials and verdicts in France in the period 1825-1833. Of special interest to Poisson was the calculation of probabilities of the two kinds of errors possible in jury verdicts, namely, the probability of convicting an innocent person and the probability of acquitting a guilty person. The U. S. Supreme Court is somewhat remiss in its decisions in ignoring these errors. They could be quite difficult to quantify in the American legal experience but some recognition of this problem could have been demonstrated.

In Williams v. Florida, the Court discusses unconditional probability of conviction for juries of sizes six and twelve, and asserts that these probabilities do not differ in any operational sense. A number of empirical studies of small group and jury behavior are referenced in Supreme Court decisions but studies of probabilistic models for jury behavior do not appear except in the Ballew decision. The probabilistic model discussed in that opinion

was subsequently shown to be weak and unwise. It may be that the results of such models are too precarious or unreliable to serve as components for Supreme Court decisions or that the law clerks and justices are uncomfortable or unfamiliar with that kind of thinking.

#### The French School: Condorcet and Laplace

Questions of this kind did perplex the French probabilists. For example, in considering the judgments of juries or tribunals, Laplace (1820) refers to the following risk principle: 'the proof of the crime of the accused ought to have a high degree of probability that the citizens have less dread of errors in judgment, if the accused be innocent and condemned, than of his new attempts and those of the unhappy ones whom the example of his impunity encourages, if he was guilty and absolved'.

This principle is attributed to Condorcet by Karl Pearson (1978). In his pioneering work on juries and testimony, Condorcet (1785) gives quite a bit of attention to the two kinds of errors inherent in a judicial decision. The discussion of conviction of an innocent defendant is given on pages 123-127 and acquitting the guilty on pages 233-241 in Condorcet's treatise. Condorcet would like the probabilities of these two kinds of errors to be quite small and he provides some development of how he would determine a probability value to be small. This also leads him to be an advocate for the abolition of capital punishment. Since capital punishment cannot be reversed, Condorcet feels that even though the probability of convicting an innocent defendant may be quite small, over a large number of cases, the probability of at least one innocent going to his or her death can be quite large. Condorcet's work on juries was motivated by questions on probabilities of judicial error and interestingly he was encouraged and supported in this work by Turgot, Controller General, one of Louis XVI's most powerful ministers.



Laplace asserts however that the probabilities of the two kinds of judicial errors are very difficult to determine. In considering this problem of errors in judicial decisions, Laplace offers implicitly the following calculations. Let  $\pi$  be the personal probability offered by a judge or juror after evaluating the evidence that a defendant is guilty. In doing this, the judge recognizes that one of the two kinds of errors may occur: erroneous convictions or erroneous acquittals. For either one, there is a cost to society or to the individual. Let us assume the loss due to erroneous conviction is  $L_C$ , similarly  $L_A$  for erroneous acquittal. These are values that can and do vary from society to society and from crime to crime. However, given these values, Laplace continues implicitly that a judge would prefer a value of  $\pi$  such that

$$\pi L_A \geq (1-\pi)L_C ,$$

that is, the expected loss due to an acquittal exceeds the expected loss due to a conviction. This leads to

$$\pi \geq \frac{L_C}{L_A + L_C} = a$$

and therefore a judge would convict when  $\pi \geq a$  (note  $0 \leq \pi \leq 1$ ).

Two judges can each have the same standard for probability of guilt, namely  $\pi \geq a$ , but of course they can differ through legal acumen as to how well they do in relative frequency of correct verdicts. Thus judges operating within this personal probability structure will have some objective frequency of success. Let  $x$  be the relative frequency of success in verdicts rendered by a judge. Laplace assumes  $1/2 \leq x \leq 1$ .

If we have  $n$  judges or a jury of  $n$  members of whom  $n-1$  convict and 1 acquit the defendant, Laplace asserts that the probability that the decision is just will be proportional to

$$x^{n-1}(1-x)^1 ,$$

likewise the probability that the opinion of the jury is not just

will be proportional to

$$(1-x)^{n-i} x^i .$$

Thus the probability of validity of the judgment of the jury is

$$\frac{x^{n-i}(1-x)^i}{x^{n-i}(1-x)^i + (1-x)^{n-i}x^i} .$$

In doing this Laplace is assuming that all the jurors are using the same threshold value for  $x$ . Laplace also assumes that the values of  $x$  are a priori equally likely to have any value between zero and one, but that  $x$  for the jurors will never be less than  $1/2$ . He further states that for any observed decision, that the jury is divided into two parts;  $n-i$  jurors vote to convict the defendant and  $i$  jurors vote to acquit and thus the relative frequency of the observed event is proportional to

$$x^{n-i}(1-x)^i + (1-x)^{n-i}x^i .$$

From before we have  $x^{n-i}(1-x)^i$  is proportional to the probability that the verdict is just and  $(1-x)^{n-i}x^i$  is proportional to the probability that the verdict is not just. Laplace has added these together and states the probability of  $n-i$  jurors voting for conviction and  $i$  jurors voting for acquittal is proportional to the sum of the two terms.

Each of these sums should be multiplied by the probability that the defendant is guilty and not guilty respectively. In the Poisson model that we discuss shortly these are taken into account explicitly as is the probability that a juror will not make an error. Laplace seems to assume that the probability of guilt and innocence, a priori, are  $1/2$  each. At the end of his analysis, Laplace defines  $P$ , the probability of a just verdict as

$$P = \frac{\int_{1/2}^1 x^{n-i}(1-x)^i dx}{\int_0^1 x^{n-i}(1-x)^i dx} .$$

Note that he integrates  $x$  from  $1/2$  to  $1$ , and that

$$\int_0^1 x^{n-1}(1-x)^i dx = \int_{1/2}^1 x^{n-1}(1-x)^i dx + \int_0^{1/2} x'^{n-1}(1-x')^i dx$$

where  $x' = 1-x$ . Note also that the combinatorial coefficient factor  $\binom{n}{i}$  is missing but for  $P$  it cancels out.

There is obviously some model inadequacy in this development and the work of Poisson and others will also bear similar frailties. Among other things, the model assumes that  $x$  is the same for each member of a jury or a panel of judges. It also assumes independence of decision by the  $n$  jurors or judges. Moreover, the distinction in judicial decision making between a panel of judges and a jury merits additional thought.

In France, at one point during Laplace's time, there were eight judges (jurors) and five determined a verdict. Essentially an initial ballot determined the outcome. This was true also shortly afterwards when seven jurors out of twelve (1825-30) and then eight jurors out of twelve (1831-33) determined the outcome. This is in contradistinction to where unanimity or something close to it is required as in the United States.

For the case of eight jurors and verdict by a vote of exactly five out of eight, the probability of an incorrect judgment is

$$1-P = \frac{\int_0^{1/2} x^5(1-x)^3 dx}{\int_0^1 x^5(1-x)^3 dx}$$

From this equation we find  $1-P = .2539$ , or a majority of one judge in a group of eight, under Laplace's model, will lead to an incorrect decision in roughly one out of four cases. In other words, we can add that the defendant's risk of being unjustly convicted or the risk of criminals escaping punishment would both be rather high. Of course, this assumes that  $x \geq 1/2$  is realistic and that Laplace's model is valid. If we now sharpen the judge's evaluation powers

and assume  $x \geq 4/5$ , we find that

$$1-P = 0.0856$$

and then the choice of an incorrect verdict is reduced to one out of twelve. This is closer to values we obtain subsequently from the Poisson model in early 19th century France and its modification and use in mid twentieth century America.

On the other hand if we still consider  $x \geq 1/2$ , require a jury of size 12 and unanimity, we find

$$1-P = \int_0^{1/2} x^{12} dx / \int_0^1 x^{12} dx = 0.0001221$$

or only one error in 8192 cases. Poisson discusses this Laplace result and shows that the probabilities of errors in convictions is 14/8192, 92/8192, 378/8192, 1093/8192, 2380/8192 when convictions are voted by 11 to 1, 10 to 2, 9 to 3, 8 to 4, and 7 to 5 respectively. Thus with the smallest majority, the probability of error is approximately 2/7, so that out of a very large number of accused convicted by a 7 to 5 majority, approximately 2/7 should not have been. For a majority conviction by 8 to 4, nearly 1/8 of the convictions could be in error. Actually, Laplace suggests that the decision rule should be at least 9 out of 12.

Poisson stresses that these results from the Laplace analysis assume the probability of guilt before trial is 1/2, an assumption he considers unrealistic and that the equation for P should read

$$P = \frac{\theta \int_{1/2}^1 x^{n-1}(1-x)^i dx}{\theta \int_{1/2}^1 x^{n-1}(1-x)^i dx + (1-\theta) \int_0^{1/2} x^{n-1}(1-x)^i dx}$$

where  $\theta$  is the probability of guilt before the accused is brought to trial. Once again the binomial factor  $\binom{n}{i}$  is not written because it cancels in the equation and, of course,  $\theta = 1/2$  gives the place referred to.

Poisson also notes that the Laplace derivation assumes that the likelihood of a juror not making an error is the same for all jurors (or the mean of a distribution of values over jurors) an assumption both Poisson and we will also grant for a specific venire but where the mean can vary for different venire. In addition, however the Laplace structure includes nothing that depends on the abilities of the jurors who render the verdict except implicitly through their estimate of the threshold value for  $a$ .

In short, we see that the Laplace analysis and its conclusions depend only on jury size and the majority producing the verdict, while Poisson asserts that  $\theta$ , the probability that a defendant is guilty before the evidence is presented, and  $\mu$ , the probability that a juror will not make an error are two additional parameters to consider in jury analyses. The value of  $\theta$  is a reflection of the society in which jury decisions are made. Poisson is quite sensitive to the fact that in a tranquil society,  $\theta > 1/2$  but that, say during the French revolution,  $\theta$  could be quite smaller than  $1/2$ . The value of  $\mu$  should depend on the characteristics of the venire from which a juror is drawn. Poisson desires that  $\mu \geq 1/2$  just as Laplace required  $x > 1/2$  for each judge in his model. At any rate, the computation of probabilities of incorrect verdicts should be based on these values.

#### Poisson Jury Model

Let us now look into the Poisson model in some detail. It is important to note that Poisson in developing this model paid heed to the data available in his day. For the period 1825-30, jury decisions were based on seven or more out of twelve jurors favoring either conviction or acquittal. Cases with verdicts of exactly seven out of twelve went to a higher court which could change the verdict. For each year, the number of trials and number of convictions were listed for crimes against persons and crimes against property. In the period 1831-33, listings were also available except the majority required was eight or more out of twelve. In

1832 and 1833, the jury could find extenuating circumstances in a conviction that would lead to a lighter penalty.

What impressed Poisson was the stability of the conviction ratios over each of the years 1825-1830 and 1832-1833. He felt this was a basis for developing a model that in some parsimonious way could reproduce the data, and if so, lead to the computation of the probabilities of the two kinds of errors important in judging the effects of size and decision making of a jury, namely, the probability of acquitting a guilty defendant. Tables 1 and 2 are taken from Poisson's work and show the stability of conviction ratios noted by him. Note that in 1832, the conviction ratio (.5388) is somewhat less than in 1832 and 1833 (.5890) even though 8 or more out of 12 are required for a conviction; extenuating circumstances leading to reduced sentences are permitted in 1832 and 1833, thus possibly serving as a factor to increase the conviction ratio. These conviction ratios are, of course, smaller than for the years 1825-1830.

To check on the homogeneity of the annual proportions of conviction over the years 1825-1830, Poisson divided the six years into two groups, 1825-1827 and 1828-1830 and tested the difference of the proportion of conviction in each period employing the normal approximation to the binomial. He concluded there was no difference. Since the  $\chi^2$  goodness of fit test over the six years is now available in our statistical armory, the homogeneity hypothesis was tested in this manner employing a  $\chi^2$  with five degrees of freedom. We computed  $\chi^2_5 = 14.04$  from the data in Table 1. Thus homogeneity is rejected at the .05 level of significance but accepted at the .01 level of significance. The principal contribution to statistical significance comes in 1830 and Poisson, in his work, remarks that possibly the proportion of convictions in that year may be a little out of line. If we omit 1830, we compute  $\chi^2_4 = 4.85$  which indicates no significance at the .05 level and thus homogeneity over the five years 1825-1829.

TABLE I

Number of Cases, Jury Decisions, and Estimates of  $\Gamma_{12,5}$  by Year in the Years 1825-1830 in France

Statistic	Crimes against	1825	1826	1827	1828	1829	1830	Total
(1) Number of accused		6652	6988	6929	7396	7373	6962	42,300
(2) Number of accused	Person	1897	1907	1911	1844	1791	1666	11,016
(3) Number of accused	Property	4755	5081	5018	5552	5582	5296	31,284
(4) Number of convicted		4037	4348	4236	4551	4475	4130	25,777
(5) Number of convicted	Person	882	967	948	871	834	766	5,268
(6) Number of convicted	Property	3155	3381	3288	3680	3641	3364	20,509
Estimates of $\Gamma_{12,5}$								
(7) Conviction ratios	Total	.6068	.6222	.6133	.6153	.6069	.5932	.6094
(8) Conviction ratios	Person	.4649	.5071	.4961	.4723	.4659	.4598	.4782
(9) Conviction ratios	Property	.6635	.6654	.6552	.6628	.6523	.6352	.6556

TABLE II

Number of Accused, Jury Decisions, and Estimates of  $\Gamma_{12,4}$  by Year  
in the Years 1831-1833 in France

Statistic	Crimes against	1831	1832	1833	1832 and 1833
(1) Number of accused		7606	7555	6964	14,519
(2) Number of accused	Person	2046	--	--	4,108
(3) Number of accused	Property	5560	--	--	10,421
(4) Number of convicted		4098	4448	4105	8,553
(5) Number of convicted	Person	743	--	--	1,889
(6) Number of convicted	Property	3355	--	--	6,665
Estimates of $\Gamma_{12,4}$					
(7) Conviction ratios	Total	.5388	.5887	.5895	.5890
(8) Conviction ratios	Person	.3631	--	--	.4598
(9) Conviction ratios	Property	.6034	--	--	.6395

From our previous discussion of Poisson's criticism of Laplace, we are aware of his concern to include two parameters;  $\theta$ , the probability that the accused is guilty before the evidence is presented to the jury and  $\mu$ , the probability that a juror will not make an error. The first parameter is a commentary on the society and its law enforcement procedures and the second relates to how well a selected juror can sift through and assess evidence. We now list  $\theta$ ,  $\mu$ , and the following definitions to develop the model.  $P_C$ : probability of a conviction,  $P_A$ : probability of an acquittal,  $P_{G/A}$ : probability of guilt given an acquittal,  $P_{I/C}$ : probability of innocence given a conviction. For an attempt of a model employing only one parameter, the reader is referred to Walbert (1971).

Subsequently when we modify the model to make it more appropriate for the American experience we will add  $P_H$ : probability of a hung jury, and instead of  $\mu$  employ  $\mu_1$ : probability that a juror will vote guilty given the accused is guilty and  $\mu_2$ : probability that a juror will vote for acquittal given the accused is innocent. A word



about  $P_{G/A}$  and  $P_{I/C}$  is in order. By guilt we mean 'convictable' and by innocence we mean 'nonconvictable' on the basis of the evidence. Only some higher being (sometimes not even the defendant) can know the true situation. Empirically the decision of a judge can be and is taken as the anchor and compared with jury decisions to estimate these errors and we will look into this later to compare the results with values obtained from our models and models of Poisson.

Since in Poisson's day the majority required for decision was first seven out of twelve and then eight out of twelve, essentially an initial ballot could suffice. Thus the probability of conviction,  $P_C$ , is the probability that say  $i$  jurors vote for acquittal where  $i \leq 5$  or  $i \leq 4$ . We can determine the probability that  $i$  jurors out of  $n$  vote for acquittal in the following way. Assume  $n = 1$ , then write

$$P_C = P_{CG} + P_{C\bar{G}}$$

where  $P_{CG}$  is the joint probability of conviction and guilt and  $P_{C\bar{G}}$  is the joint probability of conviction and innocence. Also

$$P_C = P_{C/G}P_G + P_{C/\bar{G}}P_{\bar{G}}$$

where  $P_{C/G}$  is conditional probability of conviction given guilt and  $P_{C/\bar{G}}$  is conditional probability of conviction given innocence. But  $P_{C/G} = \mu$ ,  $P_{C/\bar{G}} = 1-\mu$ , hence

$$P_C = P_G\mu + P_{\bar{G}}(1-\mu)$$

and since  $P_G = \theta$

$$P_C = \theta\mu + (1-\theta)(1-\mu)$$

or

$$P_A = \theta(1-\mu) + (1-\theta)\mu$$

but  $P_A$  is the probability that a juror will vote for acquittal.

Let  $\gamma_{n,i}$  be the probability that exactly  $i$  jurors out of  $n$  vote for acquittal; then  $\gamma_{1,1} = P_A$ . If  $n = 2$ , and  $\gamma_{2,i}$  is probability

exactly  $i$  jurors out of two vote for acquittal we have

$$\gamma_{2,0} = \theta\mu^2 + (1-\theta)\mu^2$$

$$\gamma_{2,1} = 2\theta\mu(1-\mu) + 2(1-\theta)(1-\mu)$$

$$\gamma_{2,2} = \theta(1-\mu)^2 + (1-\theta)\mu^2 .$$

Thus for a jury of size  $n$ , we get

$$\gamma_{n,i} = \binom{n}{i} [\theta\mu^{n-i}(1-\mu)^i + (1-\theta)\mu^i(1-\mu)^{n-i}] .$$

Note that the two terms in the brackets are respectively the 'guilty' component where the  $i$  votes for acquittal are in error and the 'not-guilty' component where the  $i$  votes for acquittal are not in error. Terms very similar to these two terms have appeared in the Laplace development where  $\mu$  is replaced by  $x$  and  $\theta = 1/2$ .

Define

$$\Gamma_{12,5} = \sum_{i=0}^5 \gamma_{12,i}$$

$$\Gamma_{12,4} = \sum_{i=0}^4 \gamma_{12,i}$$

and these are the probabilities of conviction when majorities required for conviction are 7 or more out of 12 and 8 or more out of twelve respectively. Estimates of  $\Gamma_{12,5}$  and  $\Gamma_{12,4}$  can be secured from the French data and thus one can produce two equations in two unknowns, namely  $\mu$  and  $\theta$ . In 1825-30, if a conviction was based on exactly seven out of twelve, another court intervened and thus the number of such cases was known by year. Since  $\Gamma_{12,4} - \Gamma_{12,5} = \gamma_{12,5}$  another anchor is provided to check on the model. The estimates of  $\gamma_{12,5}$  found in this way when checked with the empirical values reinforced the use of the model. One is faced with two equations of high degree in  $\mu$  and  $\theta$  but Poisson had some ingenious methods for making the solutions feasible.

Over all trials, Poisson obtained the estimates  $\theta = .64$ ,  $\mu = .75$ ; for crimes against persons, the estimates are  $\theta = .54$ ,  $\mu = .68$ ; for crimes against property, the estimates are  $\theta = .67$ ,  $\mu = .78$ . This demonstrates how  $\theta$  and  $\mu$  can easily vary with the criminal charge. Also, while we treat them as independent variables, this is necessarily not so, and in fact,  $\mu$  can also vary with  $n$ , the size of the jury, and  $\theta$  itself could in some societal contexts depend on  $n$ . For purposes of exposition and purposes of comparison, we will employ  $\theta = .64$  and  $\mu = .75$  since felony trials in the United States for which we have data are based on crimes against both persons and property.

Poisson is quite aware of the complementary nature of  $\theta$  and  $\mu$ , namely that  $(1-\theta)$  and  $(1-\mu)$  will produce the same probability of conviction in his model. He comments that the high proportion of convictions during the period of the French Revolution can not be employed to suggest fairness, equity, or reasonableness since values,  $\theta = .36$  and  $\mu = .25$  yield the same values for  $P_C$  as  $\theta = .64$ ,  $\theta = .75$  that were derived from his model and the data of 1825-30, 1831-33. Thus bringing to trial, an individual whose prior probability of guilt is about  $1/3$  where jurors can be in error  $3/4$  of the time, gives (in the seven or more out of twelve situations),  $P_C = .61$  just as in the case where the probability of juror error equals  $1/4$  and probability of guilt before trial is about  $2/3$ . If we assume  $\mu \geq 1/2$ ,  $\theta \geq 1/2$ , then the  $\mu$ ,  $\theta$  solutions are unique.

Armed with the results of his model, Poisson proceeds to estimate the two kinds of jury errors. For the period 1825-1830, he estimates the probability of convicting an innocent defendant is .06 (over person and property crimes) and the probability of acquitting a guilty defendant is .18. Poisson gives more results but the figures just cited will suffice to provide a basis for comparison with his 20th century successors.

To summarize, the Poisson jury model seems to serve the French jury experience quite well. There are only two parameters to produce a rather parsimonious accounting of French jury decisions in

the period 1825-1933. The data on hand, e.g. proportion of convictions by 7 or more out of 12, by 8 or more out of 12, and by exactly 7 out of 12, permit the development of two equations with two unknowns under the implicit assumption that one ballot is required. However the plurality of 7 or 8 out of 12 jurors to produce a verdict, essentially leads to only one ballot. The two parameters,  $\theta$ : the probability the defendant is guilty before the trial begins and evidence is presented, and  $\mu$ : the probability a juror will not make an error, are latent parameters. Estimates of these parameters are produced from the data on proportion of jury convictions. Poisson computed these values by solving equations of high degrees and he essentially employed the method of moments to obtain these estimates. The equations, of course, derive from

$$Y_{n,i} = \binom{n}{i} [\theta \mu^{n-i} (1-\mu)^i + (1-\theta) \mu^i (1-\mu)^{n-i}] .$$

#### The American Jury

The application of Poisson's model to the American experience requires modifications. In most felony trials unanimity is required and there are 12 lay jurors. If an initial ballot does not produce unanimity, jury deliberations take place until unanimity is achieved or there is a hopeless deadlock in achieving this goal. Therefore, in addition to  $\mu$  and  $\theta$  we require some modeling of the deliberative process leading to conviction, acquittal, or stalemate if the initial ballot does not reflect unanimity. If initial ballot American data is available, one can estimate  $\mu$  and  $\theta$  since we could consider this somewhat analogous to the French jury situation.

While much jury data may exist in raw form in many state and federal archives, there is only one published account of initial ballot results and final decisions of some juries. This data appears in Kalven and Zeisel (1966). For 225 juries, each of size 12, there is reported the votes on the initial ballots and the juries' final decisions. Table 3 lists this data. It will permit us to obtain estimates of  $\mu$  and  $\theta$  from initial ballot data for the American

TABLE III

Distribution of Votes for Acquittal on First Ballot and Jury Decisions

Number of Votes for Acquittal on First Ballot

0	1 - 5	6	7 - 11	12	Total
0 0%	5 5%	5 50%	37 91%	26 100%	73 32%
43 100%	90 86%	5 50%	1 2%	0 0%	139 62%
0 0%	10 9%	0 0%	3 7%	0 0%	13 6%
43 100%	105 100%	10 100%	41 100%	26 100%	225 100%
19	47	4	18	12	100%

Not Guilty

Guilty

Hung

Total

Final  
Verdict

Percent of  
Total of 225  
Cases

scene. In fact, we will consider  $\mu_1$ : probability the juror will vote guilty given the defendant is guilty and  $\mu_2$ : probability the juror will vote not guilty given the defendant is innocent, that is,  $\mu$ : probability the juror will not make an error, is sharpened. We will then require some modeling to take us from the initial ballot to final decision and the data in Table 3 will also be helpful in this regard, although we jump from initial ballot to final decision in one step. In a subsequent section we will discuss going from initial ballot to final decision in several steps but will be hampered by the lack of data on what goes on in the American jury room.

To include  $\mu_1$  and  $\mu_2$  in  $\gamma_{n,i}$ , we can slightly revise  $\gamma_{n,i}$  as it appeared in the previous section to obtain

$$\gamma'_{n,i} = \binom{n}{i} [\theta \mu_1^{n-i} (1-\mu_1)^i + (1-\theta) \mu_2^i (1-\mu_2)^{n-i}]$$

where  $\gamma_{n,i}$  is the probability that a jury of size  $n$  casts  $i$  votes for acquittal on the first ballot. We also have  $\Gamma_{n,i} = \sum_{j=1}^i \gamma_{n,j}$  where  $\Gamma_{n,i}$  is probability of at most  $i$  votes for acquittal on first ballot and we can define

$$p_{n,i} = \frac{\binom{n}{i} \theta \mu_1^{n-i} (1-\mu_1)^i}{\gamma_{n,i}}$$

where  $p_{n,i}$  is the probability that the accused is guilty given exactly  $i$  votes for acquittal on the first ballot. Likewise

$P_{n,i} = \sum_{j=0}^i p_{n,j} \gamma_{n,j} / \Gamma_{n,i}$  where  $P_{n,i}$  is the probability that the accused is guilty given at most  $i$  votes for acquittal on the first ballot.

#### Other Estimation Approaches and Extension of the Model

Let us consider Table 3 in the following way. We can think of the first-ballot results for the 225 trials as 225 independent observations from a five-cell multinomial distribution. Under the two-parameter model the cell probabilities are  $p_1 = \gamma_{12,0}$ ,  $p_2 = \sum_{i=1}^5 \gamma_{12,i}$ ,  $p_3 = \gamma_{12,6}$ ,  $p_4 = \sum_{i=7}^{11} \gamma_{12,i}$ , and  $p_5 = \gamma_{12,12}$

respectively, while under the three-parameter model the cell probabilities denoted by primes are identical to the above with  $\gamma'$  replacing  $\gamma$ . Two estimation approaches arise naturally from such a basis, the method of maximum likelihood and the method of modified minimum  $\chi^2$ . In either case and under either model we restrict ourselves to solutions for the parameters on the interval  $(1/2, 1)$  since it is difficult to believe  $\mu < 1/2$ ,  $\theta < 1/2$  in American society.

In the case of the maximum likelihood estimators we examine

$$L(\mu, \theta) = c(p_1)^{43} (p_2)^{105} (p_3)^{10} (p_4)^{41} (p_5)^{26}$$

over the range  $1/2 \leq \mu < 1$ ,  $1/2 \leq \theta < 1$  where the exponents are the total number of verdicts for each acquittal vote category; and

$$L(\mu_1, \mu_2, \theta) = c(p'_1)^{43} (p'_2)^{105} (p'_3)^{10} (p'_4)^{41} (p'_5)^{26}$$

over the range  $1/2 \leq \mu_1 < 1$ ,  $1/2 \leq \mu_2 < 1$ ,  $1/2 \leq \theta < 1$ . The unique solutions are  $\mu = .88$ ,  $\theta = .69$  and  $\mu_1 = .86$ ,  $\mu_2 = .92$ ,  $\theta = .70$ , respectively, Gelfand and Solomon (1977).

As for the modified minimum  $\chi^2$  estimators we minimize

$$\chi^2(\mu, \theta) = \sum_{i=1}^5 \frac{(O_i - 225p_i)^2}{225p_i}$$

and

$$\chi^2(\mu_1, \mu_2, \theta) = \sum_{i=1}^5 \frac{(O_i - 225p'_i)^2}{225p'_i}$$

where  $O_1 = 43$ ,  $O_2 = 105$ ,  $O_3 = 10$ ,  $O_4 = 41$  and  $O_5 = 26$  and the ranges of the parameters are restricted as above. The unique solutions are  $\mu = .84$ ,  $\theta = .66$  and  $\mu_1 = .92$ ,  $\mu_2 = .92$ ,  $\theta = .76$ , respectively. The results of these two estimation procedures along with estimates by the method of moments are displayed in Table 4 and indicate reasonably good agreement.

Consideration of the situation in terms of a multinomial distribution opens the possibility of a wide variety of extensions of

TABLE IV

## Estimates of the Model Parameters

	2-Parameter Model	3-Parameter Model
Case I: Modified Minimum $\chi^2$	$\mu = .84 \quad \theta = .66$	$\mu_1 = .92$ $\mu_2 = .92 \quad \theta = .76$
Case II: Maximum Likelihood	$\mu = .88 \quad \theta = .69$	$\mu_1 = .86$ $\mu_2 = .92 \quad \theta = .70$
Case III: Method of Moments	$\mu = .90 \quad \theta = .70$	$\mu_1 = .90$ $\mu_2 = .92 \quad \theta = .66$

of the basic model limited only by the availability of data. For example, starting with the two-parameter model one might instead wish to think of juries composed of fixed numbers of men,  $n_m$  and of women,  $n_w$  ( $n_m + n_w = n$ ) or perhaps of juries composed of  $n_b$  blacks and  $n_w$  whites ( $n_b + n_w = n$ ). Instead of a common  $\mu$  for all jurors, associate a  $\mu_m$  and  $\mu_w$  to male and female jurors respectively (similarly for blacks and whites). The effect of such additional parametrization leads to consideration of  $I$ , the number of first-ballot votes for acquittal, as the sum of two independent random variables (i.e., the number of male votes for acquittal plus the number of female votes for acquittal and similarly for blacks and whites). Thus the distribution of  $I$  results from a convolution, i.e., in the male-female case

$$P(I=i) = \gamma_{n,i}(\mu_m, \mu_w, \theta) = \sum_{j=\max(0, i-n_w)}^{\min(n_m, i)} \gamma_{n_m, j}(\mu_m, \theta) \gamma_{n_w, i-j}(\mu_w, \theta).$$

From this example it is obvious that additional complexity can be inserted into the model and that the three-parameter model can also be extended similarly. The number and definition of the



multinomial cells is flexible. Hence with appropriately gathered first-ballot data and effective computer programs the parametric estimation possibilities are quite broad. At the present time the prospects for availability of such data as described above are 'at best' slim.

In examining the distribution of  $I$ , the number of first-ballot votes for acquittal, we notice that under either the two- or three-parameter model we find it to be a mixture of binomials. In fact under the two-parameter model

$$P(I=i) = \gamma_{n,i} = \theta P[I=i | I \sim B_1(n, 1-\mu)] + (1-\theta) P[I=i | I \sim B_1(n, \mu)]$$

and under the three-parameter model

$$P(I=i) = \gamma'_{n,i} = \theta P[I=i | I \sim B_1(n, 1-\mu_1)] + (1-\theta) P[I=i | I \sim B_1(n, \mu_2)] ,$$

where  $B(n, p)$  is the binomial with parameters  $n$  and  $p$ . Apart from the earlier discussion leading to these models, such a mixture is ideal for describing the expected bimodal distribution of first-ballot votes. We have  $E(I) = n[\theta(1-\mu) + \mu(1-\theta)]$  and  $n[\theta(1-\mu_1) + \mu_2(1-\theta)]$  respectively, while  $\text{Var}(I) = n\mu(1-\mu)$  and  $n[\theta\mu_1(1-\mu_1) + (1-\theta)\mu_2(1-\mu_2)]$  respectively. Under the range of values for the parameters suggested by Table 4 and under either model with  $n = 12$ ,  $E(I)$  is approximately equal to four and  $\text{Var}(I)$  is approximately equal to one. Since Table 4 suggests little difference between  $\mu_1$  and  $\mu_2$ , we shall use the two-parameter model  $(\theta, \mu)$  and its estimates for the remainder of this exposition.

The values for  $\theta$  and  $\mu$  that we have computed from the Kalven-Zeisel data result from initial ballot responses for American juries of size 12. Since juries in Poisson's day essentially were, or could be conceived of, as one ballot juries; 19th century French and 20th century American values for  $\theta$  and  $\mu$  may be commensurate. The American  $\theta$  is a bit higher and the American  $\mu$  is quite a bit higher than their French counterparts. Assuming this conclusion is valid, it would be interesting to examine the difference in the  $\mu$ 's - e.g. is the American juror more sophisticated? As for  $\theta$ , is

the search and interrogation process and the public climate on crime in America doing a better job in bringing miscreants to trial. This can be somewhat misleading because i) the difference in  $\theta$ 's is not great, and ii) defendants who go to jury trial are the very few for whom plea bargaining has not been successful, the crime is serious and the evidence of guilt is not ironclad.

In order to provide fully for an American jury model, it is necessary to allow for jury deliberation if the initial ballot does not yield unanimity for conviction or acquittal. Our search for conditional probabilities of conviction given innocence and acquittal given guilt must rely on an explication of the full model. From the Kalven-Zeisel data we note that 'majority persuasion' is taking place in bringing the jury from initial ballot to final verdict. The number of votes for acquittal on the initial ballot seems to determine the outcome except for some infrequent reversals. For example, for 1 to 5 votes for acquittal on the initial ballot, only five percent of the time is there a final verdict of not guilty; for 7 to 11 votes for acquittal on the initial ballot, only two percent of the time is there a final guilty verdict. In the former situation there is also a nine percent chance of a hung jury and for the latter situation, there is a seven percent chance for a hung jury

Let us suppose that the first ballot majority always prevails and in the case of an evenly split first ballot there is an even chance which way the decision will go. The latter assumption is borne out by the data in Table 3 but there only ten times out of 225 cases when the vote is evenly split. Under such assumptions, the probability of conviction,  $P_C$  is

$$P_C = \sum_{i=0}^5 \gamma_{12,i} + \frac{1}{2} \gamma_{12,6}$$

with  $P_A = 1 - P_C$  and  $P_H$ , the probability of a hung jury, equal to zero.

Then the probability that a defendant is guilty given conviction,  $P_{G|C}$  is

$$P_{G|C} = \frac{\sum_{i=0}^5 P_{12,i} \gamma_{12,i} + \frac{1}{2} P_{12,6} \gamma_{12,6}}{P_C}$$

and  $P_{I|C} = 1 - P_{G|C}$ . Likewise, the probability that a defendant is innocent given acquittal,  $P_{I|A}$ , is

$$P_{I|A} = \frac{\sum_{i=7}^{12} (1 - p_{12,i}) \gamma_{12,i} + \frac{1}{2} P_{12,6} \gamma_{12,6}}{P_A}$$

and  $P_{G|A} = 1 - P_{I|A}$ .

Employing the maximum likelihood values,  $\theta = .69$ ,  $\mu = .88$ , Gelfand and Solomon (1975) show that the conditional quantities  $P_{G|A}$  and  $P_{I|C}$  are approximately twenty times larger for a jury of size six than for a jury of size twelve. Naturally this assumes the values of  $\theta$  and  $\mu$  remain the same for the two jury sizes. Also this approximation by simple majority persuasion is somewhat crude and we now seek to better this model.

A next step in refining this model is to modify the pure majority persuasion aspect by the Kalven-Zeisel data. For example we can write

$$P_C = \gamma_{12,0} + (.86) \sum_{i=1}^5 \gamma_{12,i} + .5 \gamma_{12,6} + .02 \sum_{i=7}^{11} \gamma_{12,i}$$

since in 86 percent of the trials where the initial ballot had 1 to 5 votes for acquittal the final decision was a guilty verdict and in two percent where the initial ballot had 7 to 11 votes for acquittal, a guilty verdict was rendered.

We can also write

$$P_A = \gamma_{12,12} + .91 \sum_{i=7}^{11} \gamma_{12,i} + .5 \gamma_{6,6} + 0.5 \sum_{i=1}^5 \gamma_{12,i}$$

and then

$$P_H = 1 - P_A - P_C$$

obtaining for the first time a value for  $P_H$  that is not equal to zero. These also lead to values for  $P_{G|C}$  and  $P_{I|A}$  and consequently  $P_{I|C} = 1 - P_{G|C}$  and  $P_{G|A} = 1 - P_{I|A}$ . Once again these conditional probabilities are shown to vary considerably, Gelfand and Solomon (1975), between juries of size six and size twelve.

For our final modification of the model in going from initial ballot to final verdict we employ a blend of theory and empirical evidence from mock jury data. In a subsequent section we try our hand at taking the jury through several ballots from initial decision to final verdict. For our mock jury data, we use the results of studies conducted by Davis and his collaborators who develop social decision schemes to take the jury from initial ballot directly to final verdict. One such scheme we employ and then modify is by Davis (1973):

Votes for Acquittal on First Ballot

	1	0	1	2	3	4	5	6	7	8	9	10	11	12
Prob.	$P_C(i)$	1	1	$\frac{20}{24}$	$\frac{18}{24}$	$\frac{16}{24}$	$\frac{14}{24}$	$\frac{12}{24}$	$\frac{5}{24}$	$\frac{4}{24}$	$\frac{3}{24}$	$\frac{2}{24}$	0	0
of	$P_A(i)$	0	0	$\frac{2}{24}$	$\frac{3}{24}$	$\frac{4}{24}$	$\frac{5}{24}$	$\frac{6}{24}$	$\frac{14}{24}$	$\frac{16}{24}$	$\frac{18}{24}$	$\frac{20}{24}$	1	1
Decision	$P_H(i)$	0	0	$\frac{2}{24}$	$\frac{3}{24}$	$\frac{4}{24}$	$\frac{5}{24}$	$\frac{6}{24}$	$\frac{5}{24}$	$\frac{4}{24}$	$\frac{3}{24}$	$\frac{2}{24}$	0	0

Note that in the Kalven-Zeisel data of Table 3 we are restricted to five columns because the initial ballot data has been aggregated that way. Here through the experimental study we have all thirteen columns for acquittal votes on first ballot. Note also that majority persuasion is exhibited by the results of the experimentation.

To incorporate this data in a meaningful way we return to our modeling. We first consider the twelve member jury. Given a first

ballot stance, i.e. number of votes for acquittal on the first ballot, we wish the probabilities associated with each of the three possible jury conclusions. Label these three probabilities as  $P_C(i)$ ,  $P_A(i)$ ,  $P_H(i)$  where, for instance,  $P_C(i)$  is the probability the jury ultimately convicts given  $i$  votes for acquittal on the first ballot. Obviously for any fixed value of  $i$ , the sum of the three probabilities is one, and although each juror has two choices, the jury has three. Employing our previous notation, we write

$$P_C = \sum_{i=0}^{12} P_C(i) \gamma_{12,i}$$

$$P_A = \sum_{i=0}^{12} P_A(i) \gamma_{12,i}$$

$$P_H = \sum_{i=1}^{12} P_H(i) \gamma_{12,i}$$

or the matrix equation

$$\underline{P} = \underline{D}\underline{Y}$$

where  $\underline{P}$  is a column vector with three rows or equivalently

$\underline{P}' = (P_C, P_A, P_H)$ ,  $\underline{Y}$  is a column vector with 13 rows or

$\underline{Y}' = (\gamma_{12,0}, \gamma_{12,1}, \dots, \gamma_{12,12})$ , and  $\underline{D}$  is a social decision matrix with three rows and thirteen columns.

If we now evaluate  $\gamma_{12,i}$  for all  $i$  using  $\theta = .69$ ,  $\mu = .88$ , and employ the social decision scheme from Davis, we can compute  $\underline{P}$  and then compare the three coordinates with empirical values from Kalven-Zeisel. Kalven-Zeisel's study included 3576 jury trials (for only 225 were initial ballots known) and for these we have  $\hat{\underline{P}}' = (.642, .303, .055)$ . In fact for the 225 trials we have  $\hat{\underline{P}}' = (.62, .32, .06)$ . Both  $\underline{P}'$  estimates are from the data.

When the  $\underline{P}'$  coordinates are obtained from the model, the resultant goodness-of-fit statistic is  $\chi^2 = 8.62$  which is acceptable at the .01 level. By some very slight adjustments in the matrix  $\underline{D}$  supplied by Davis (1973), Gelfand and Solomon (1977) achieve

Number of Votes for Acquittal on Initial Ballot

	1	0	1	2	3	4	5	6	7	8	9	10	11	12
Prob.	$P_C(i)$	1	1	$\frac{20}{24}$	$\frac{18}{24}$	$\frac{16}{24}$	$\frac{14}{24}$	$\frac{12}{24}$	$\frac{5}{24}$	$\frac{4}{24}$	$\frac{2}{24}$	$\frac{1}{24}$	0	0
of	$P_A(i)$	0	0	$\frac{1}{24}$	$\frac{2}{24}$	$\frac{4}{24}$	$\frac{5}{24}$	$\frac{6}{24}$	$\frac{14}{24}$	$\frac{16}{24}$	$\frac{18}{24}$	$\frac{20}{24}$	1	1
Decision	$P_H(i)$	0	0	$\frac{3}{24}$	$\frac{4}{24}$	$\frac{4}{24}$	$\frac{5}{24}$	$\frac{6}{24}$	$\frac{5}{24}$	$\frac{4}{24}$	$\frac{3}{24}$	$\frac{3}{24}$	1	1

and we obtain  $\hat{\hat{P}}' = (.6419, .3024, .0557)$  leading to  $\chi^2 = .0841$  (virtually a perfect fit);  $\hat{\hat{P}}'$  are estimates from the model.

Values that depart from  $\theta = .69$ ,  $\mu = .88$  yield a model that does not reproduce the observed  $P$  vector. This suggests that we can now, within this model, attempt the conditional probabilities of interest

$$P_{G|C} = \frac{\sum_{i=0}^{12} P_C(i) p_{12,i} \gamma_{12,i}}{\sum_{i=0}^{12} P_C(i) \gamma_{12,i}}$$

$$P_{I|A} = \frac{\sum_{i=0}^{12} P_A(i) (1-p_{12,i}) \gamma_{12,i}}{\sum_{i=0}^{12} P_A(i) \gamma_{12,i}}$$

where  $P_{I|C} = 1 - P_{G|C}$  and  $P_{G|A} = 1 - P_{I|A}$ . We now get, after some computer calculations,

$$P_{I|C} = .0221$$

$$P_{G|A} = .0615 .$$

If we now try this approach for six member juries there are additional problems mainly because of lack of data. However we now attempt  $P_{I|C}$  and  $P_{G|A}$  for this situation. First we assume  $\theta$  and  $\mu$  are the same and this can easily be challenged. Then we propose a

social decision scheme for six member juries embodying majority persuasion but the cell entries we choose can also be easily contested. However, the aforementioned work of Davis and his collaborators has also provided similar six member jury decision schemes from mock jury experiments. Employing the social decision matrix given below

		Number of Votes for Acquittal on Initial Ballot							
		1	0	1	2	3	4	5	6
Probability of Decision	$P_C(i)$	1	$\frac{10}{12}$	$\frac{8}{12}$	$\frac{6}{12}$	$\frac{2}{12}$	$\frac{1}{12}$		0
	$P_A(i)$	0	$\frac{1}{12}$	$\frac{2}{12}$	$\frac{3}{12}$	$\frac{8}{12}$	$\frac{10}{12}$		1
	$P_H(i)$	0	$\frac{1}{12}$	$\frac{2}{12}$	$\frac{3}{12}$	$\frac{2}{12}$	$\frac{1}{12}$		0

we obtain  $P_C = .6347$ ,  $P_A = .3207$ ,  $P_H = .0446$ . While these values are very close to what experience shows with 12 member juries, it may be quite different for six member juries. On the other hand, one may feel quite comfortable with replicating what has been going on in society. In either event, if we continue we obtain

$P_{I|C} = .0325$  and  $P_{G|A} = .1395$ . This demonstrates quite crucial differences; a six member jury will convict 50% more innocent defendants and will set free twice as many guilty defendants. Naturally these results rely on the assumptions of the model and the employment of the Kalven-Zeisel data. It is interesting to recall Poisson's estimates based on twelve member juries and majority decisions a century and a half ago in France; namely  $P_{I|C} = .06$  and  $P_{G|A} = .18$ .

It may also be instructive to look at some empirical results. Baldwin and McConville (1979) in their book Jury Trials report on twelve member jury decisions during 1975-76 in Birmingham, England, and estimate the two conditional probabilities of error. The anchor here as to what the 'true' situation (convictable, non-convictable) might be is the judges' assessment as well as the view of police,

prosecution and defense attorneys. These are then compared with the jury decision and lead to  $P_{I|C} \geq .05$  and  $P_{G|A} > .36$ . The value of  $P_{G|A}$  seems quite high. Kalven and Zeisel also present tables of judge and jury disagreement. From one of their tables (Table 12, p. 58), we see that over 3,576 trials the jury convicts 3 percent of the time when the judge would acquit and the judge convicts 19 percent of the time when the jury acquits. In this analysis hung jury trials are omitted. The estimates of  $P_{I|C} = .03$  and  $P_{G|A} = .19$  (assuming the judge's verdict is the truth) may be contrasted with the other values just quoted. As in the British empirical experience, the estimate of  $P_{G|A}$  seems rather high.

The Poisson model we have modified by a majority persuasion decision scheme to go from initial to final ballot will yield smaller values for  $P_{I|C}$  and  $P_{G|A}$  as the jury size increases. This suggests that the traditional jury of size twelve be increased in number to lessen the two risks. This could be more costly - the major argument for juries of size six instead of twelve revolves around cost - but other issues not yet treated would require consideration. Studies of group behavior indicate that dead time and poorer performance could result from increases in group size.

There is some literature on this phenomenon but more research would be required on juries of size 12 to 24 (roughly the grand jury size in the U.S.) to study the inhibiting effects, if any, of large jury sizes. For additional discussion of mathematical models of jury decision making, Grofman (1981) presents the state of the art up to the present. His paper gives a detailed account of various models including the Gelfand-Solomon modification of the Poisson development featured here.

We have already remarked on the availability of jury data. For the past dozen years, data on number of jury trials and jury decisions by crime in the U.S. Federal Courts have been published annually. A further breakdown would, of course, be helpful. What appears annually now is exactly the kind of information available



to Poisson except that the categories of crime enumerated may be more numerous. For those who wish to develop and test models we do not seem better off than Poisson in connection with empirical data although a body of mock jury data is growing.

#### The Jury from Within: Markovian Models

If we wish to consider the American jury from initial ballot to final verdict by allowing additional balloting we will not have an empirical base. In what follows we nevertheless try our hand at this, keeping in mind the liabilities thus imposed. Klevorick and Rothschild (1979) develop another multi-ballot model and stress caution, as we do, because of the simplifying assumptions employed. In Penrod and Hastie (1979), there is some discussion of multi-ballot models. Much of what follows appears in an unpublished report by Gelfand and Solomon (1974).

Let us consider from ballot to ballot how the jury ultimately arrives at a decision. Given the paucity of data on behavior in the jury room, this presents a formidable estimation problem because of the increase in the number of parameters in such a model. From a stochastic point of view, we wish to develop both stationary (homogeneous) and non-stationary (non-homogeneous) Markov models with appropriate transition matrices, whose entries give the probability of  $j$  votes for acquittal on the  $n+1^{\text{st}}$  ballot given  $i$  votes for acquittal on the  $n^{\text{th}}$  ballot. Let us denote this probability by  $p_n(j|i)$ . The Markovian assumption seems quite reasonable, but an assumption of stationarity most likely is not. This extended model can be modified to obtain fewer states by grouping ballot outcomes, thus requiring estimation of fewer parameters. Ultimately we shall do this, but for now, regardless of the number of states, it is critical to note that the states where  $i = 0$  (i.e., all guilty votes) and  $i = 12$  (i.e., all non-guilty votes) are absorbing barriers. Moreover the chain has a finite state space, and it is reasonable to postulate that for any intermediate state an absorbing state is accessible. Hence the assumption of a stationary transition matrix suggests all other states must be transient. This leads to the

rather unsatisfactory assumption that all juries arrive at either a guilty or innocent verdict, i.e., there is no possibility of a hung jury and so some modification is required to establish this.

Actually we are overstating the situation in the sense that the remarks above imply that the probability of remaining in a 'transient' state goes to zero as the number of ballots,  $n$ , goes to  $\infty$ . Since for finite  $n$  there will be positive probability of not being absorbed as yet (i.e., 'hanging'), perhaps it is just a question of deciding on a finite number of ballots for the jury to achieve a decision or declare themselves deadlocked. That is, given enough ballots, unanimity would be reached or we could insure that say, 5% of the time the jury is hung. For majority verdicts, the percentage of hung juries should be less than 5%.

However, one could make a stronger argument for a nonstationary structure as follows. On the early ballots the number of votes for acquittal,  $i$ , may change quite a bit, but after a few ballots the jurors begin to 'lock into a position' after which the jury stance will change perhaps a vote or not at all. Thus the transition matrix cannot be stationary but in fact should be tending to an identity matrix, i.e., all states ultimately becoming absorbing. Therefore even given an infinite number of ballots the jury would not necessarily achieve a unanimous position. Since this is somewhat tentative and exploratory, we shall examine both models.

Let us first consider a stationary Markov setting. We describe a transition matrix  $P = \{p_{ij}\}$  where  $p_{ij} \equiv p_n(j|i) = p(j|i)$  (i.e., is independent of  $n$ ). Just as we have done for the initial distribution of votes for acquittal we shall describe the conditional distribution of votes for acquittal on the present ballot given  $i$  votes for acquittal on the previous ballot by a mixed binomial where the parameter values depend on  $i$ . That is, for  $i = 0, 1, \dots, 12$ ,  $p_{ij} = \gamma'_{12,j}[\mu_1(i), \mu_2(i), \theta(i)]$ ,  $j = 0, 1, \dots, 12$ . Actually we will set  $\mu_1(i) = \mu_2(i) = \mu(i)$  for convenience so that  $p_{ij} = \gamma'_{12,j}[\mu(i), \theta(i)]$ . We note that these assumptions characterize each row of  $P$  by two unknown parameters instead of requiring

13 (actually 12) parameters, thereby simplifying matters considerably. As long as a stationary transition structure is assumed such conditional distributions for appropriate  $\mu(i)$  and  $\theta(i)$  seem satisfactory.

Although no actual data is available on the average number of ballots to achieve a hung jury, Kalven and Zeisel (1966, pp. 458-459) report that in nearly 80% of trials resulting in hung juries the deliberation time is between two and ten hours. Thus we might guess at least five to at most 20 ballots will be taken with perhaps a median around ten. Naturally this is purely speculative. Our goal will be to select  $\mu(i)$  and  $\theta(i)$  such that  $\Pi \equiv (P)^n$  for  $n$  approximately 10 will be a good approximation to  $D$  as suggested and modified in Gelfand and Solomon (1975, 1977) respectively. In other words the first column vector of  $\Pi$  should approximate the first row of  $D$ , the last column vector of  $\Pi$  should approximate the second row of  $D$  and the sum of the remaining column vectors of  $\Pi$  would yield a vector that approximates the last row of  $D$ .

In selecting  $\mu(i)$  and  $\theta(i)$  we first observe that Table 5 suggests suitable choices for  $\theta(i)$  are

$$\begin{aligned}\theta(i) &= 1 & i &= 0, \dots, 5 \\ \theta(i) &= 0 & i &= 7, \dots, 12 \\ \theta(i) &= 2/3 & i &= 6.\end{aligned}$$

This follows since the value  $p_{12,i}$  is effectively the appropriate choice of  $\theta$  to use given  $i$  votes for acquittal on the first ballot. Moreover given the first-ballot position, the transition distribution would probably be unimodal, i.e.,  $\theta(i) = 0$  or  $1$ . Due to the essentially symmetric structure of  $D$  we set  $\mu(i) = 1 - \mu(12-i)$  for all  $i$ , with  $\mu(6) = 1/2$  and then experiment over various choices of  $\mu(i)$ ,  $i = 1, \dots, 5$ . A rather satisfactory fit was achieved for  $\mu(i) = 1 - (.08)i + .02$  as indicated by Table 6. Only at  $i = 6$  is the fit poor, and as observed earlier the effect will be insignificant. The fit can be somewhat refined but to no particular advantage. In

TABLE V

## A More Detailed Distribution of First-Ballot Votes

Case I		$\mu = .84$		$\theta = .66$	
	$\gamma_{12,i}$	"guilty" part	"innocent" part	$P_{12,i}$	$P_{12,i}$
0	.0814	.0814	0	1	1
1	.1862	.1862	0	1	1
2	.1950	.1950	0	1	1
3	.1238	.1238	0	1	1
4	.0531	.0531	0	1	1
5	.0165	.0165	0	1	1
6	.0055	.0036	.0019	.6545	.9971
7	.0089	.0006	.0083	.0674	.9849
8	.0274	.0001	.0273	.0016	.8993
9	.0638	0	.0274	0	.8670
10	.1005	0	.1005	0	.7660
11	.0959	0	.0959	0	.6892
12	.0420	0	.0420	0	.6600

Case II		$\mu = .88$		$\theta = .69$	
	$\gamma_{12,i}$	"guilty" part	"innocent" part	$P_{12,i}$	$P_{12,i}$
0	.1488	.1488	0	1	1
1	.2435	.2435	0	1	1
2	.1826	.1826	0	1	1
3	.0830	.0830	0	1	1
4	.0255	.0255	0	1	1
5	.0056	.0055	.0001	.9821	.9999
6	.0013	.0009	.0004	.6923	.9993
7	.0026	.0001	.0025	.0385	.9957
8	.0115	0	.0115	0	.9837
9	.0373	0	.0373	0	.9302
10	.0820	0	.0820	0	.8376
11	.1094	0	.1094	0	.7394
12	.0669	0	.0669	0	.6900

TABLE V (Cont'd)

Case III		$\mu = .9$		$\theta = .7$	
	$Y_{12,i}$	"guilty" part	"innocent" part	$P_{12,i}$	$P_{12,i}$
0	.1977	.1977	0	1	1
1	.2636	.2636	0	1	1
2	.1611	.1611	0	1	1
3	.0597	.0597	0	1	1
4	.0149	.0149	0	1	1
5	.0027	.0027	0	1	1
6	.0005	.0003	.0002	.60	.9997
7	.0012	0	.0012	0	.9980
8	.0064	0	.0064	0	.9890
9	.0256	0	.0256	0	.9545
10	.0690	0	.0690	0	.8724
11	.1130	0	.1130	0	.7647
12	.0847	0	.0847	0	.7000

any case the modeling formulation should be clear. The approximation displayed in the lower half of Table 6 can itself be taken as a social decision scheme and thus considered in terms of how well it fits the Kalven and Zeisel data in a manner analogous to previously examined schemes. In particular for Case II of Table 5 we obtain an expected distribution vector  $\tilde{P}^T = (.6355, .2976, .0669)$  which when compared with the observed vector  $\hat{P}^T$  yields a  $\chi^2$  value of 6.71. This is remarkably small in view of the crudeness of our assumptions.

In examining a nonstationary Markovian approach the technique just described can not work, for if  $\mu(i)$  and  $\theta(i)$  are now allowed to depend on  $n$ , there will be no way to select them such that  $p_n(i|i)$  increases to 1 as  $n$  increases, i.e., such that the conditional distributions at each  $i$  value will tend to a degeneracy at that value. More elaborate specification will be required for each  $i$ , considerably complicating the situation. Thus we shall examine instead the collapsed model described by Table 3 where the number of states is reduced to five. Specifically we will define the states as  $S_1: i = 0$ ,  $S_2: i = 1, 2, 3, 4, 5$ ,  $S_3: i = 6$ ,

$S_4: i = 7, 8, 9, 10, 11$  and  $S_5: i = 12$ . Note that this interpretation is forced upon us by Table 6. Otherwise we might choose  $S_2: i = 1, 2, 3, 4$ ,  $S_3: i = 4, 5, 6$  and  $S_4: i = 8, 9, 10, 11$  to better separate the states. Let us denote the transition matrix from the  $n^{\text{th}}$  to the  $n+1^{\text{st}}$  ballot by  $Q_n$ , a five-by-five matrix where, in analogy with the larger situation, we have  $q_n(j|i)$  as the entries in  $Q_n$ . At this point, if stationarity were assumed, then as in the 13 state model, we get  $Q_n = Q$  and  $q_n(j|i) = q(j|i) = q_{ij}$ . A reasonable form for  $Q$  is

$$Q = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ q(1|2) & q(2|2) & q(3|2) & q(4|2) & 0 \\ 0 & q(2|3) & q(3|3) & q(4|3) & 0 \\ 0 & q(2|4) & q(3|4) & q(4|4) & q(5|4) \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Specifying three elements in Row 4, two elements Row 3, and three elements in Row 4 we can proceed exactly as before in an effort to approximate the decision pattern contained in Table 3. Again we can achieve a satisfactory fit.

Let us now turn to the nonstationary case. In particular, let us suppose the sort of non-homogeneous behavior described earlier; that is, as a result of the first few ballots there may be a considerable change in the jury stance, but then the situation stabilizes and at most one juror will change his vote. Hence we assume

- (i)  $q_n(1|1) = 1, q_n(i|1) = 0, i = 2, 3, 4, 5$  for all  $n$   
 $q_n(5|5) = 1, q_n(i|5) = 0, i = 1, 2, 3, 4$  for all  $n$
- (ii)  $q_n(1|2) = 0, i = 3, 4, 5$  for  $n \geq 3$   
 $q_n(i|4) = 0, i = 1, 2, 3$  for  $n \geq 3$
- (iii)  $q_n(i|3) = 0, i = 1, 3, 5$  for all  $n$ .

If in assumption (ii)  $n \geq 3$  seems too early for such stability, it is easy to adjust what follows for a bit larger  $n$ . Assumption

TABLE VI  
A Comparison of Two Social Decision Schemes

I. As Suggested by Davis and Modified by the Authors

	1	2	3	4	5	6	7	8	9	10	11	12
$P_C(i)$	1	1	.83	.75	.66	.58	.50	.21	.17	.12	.08	0
$P_A(i)$	0	0	.04	.08	.17	.21	.50	.58	.66	.75	.83	1
$P_H(i)$	0	0	.13	.17	.17	.21	0	.21	.17	.12	.08	0

II. Obtained via a Stationary Markov Structure Assumption

	1	2	3	4	5	6	7	8	9	10	11	12
$P_C(i)$	1	.96	.88	.78	.68	.56	.37	.20	.12	.06	.03	0
$P_A(i)$	0	.00	.03	.06	.12	.20	.37	.56	.68	.78	.88	1
$P_H(i)$	0	.04	.09	.16	.20	.24	.26	.24	.20	.16	.09	0

(iii) becomes implicit beyond such an  $n$  as specified by (ii) and is not unreasonable nor critical for smaller  $n$  in view of the infrequency of occurrence of  $S_3$  and the initial instability. Again it can be modified if desired. Let us define

$$Q(C|2) = P(\text{conviction given state } S_2 \text{ on first ballot})$$

$$Q(A|2) = P(\text{acquittal given state } S_2 \text{ on first ballot})$$

$$Q(C|4) = P(\text{conviction given state } S_4 \text{ on first ballot})$$

$$Q(A|4) = P(\text{acquittal given state } S_4 \text{ on first ballot})$$

Under (i), (ii), (iii) we determine

$$Q(C|2) = q_1(1|2) + q_1(2|2)q_2(1|2) + U[q_1(3|2)q_2(2|3) + q_1(2|2)q_2(2|2)]$$

$$Q(A|2) = Vq_1(3|2)q_2(4|3)$$

$$Q(C|4) = Uq_1(3|4)q_2(2|3)$$

$$Q(A|4) = q_1(5|4) + q_1(4|4)q_2(5|4) + V[q_1(3|4)q_2(4|3) + q_1(4|4)q_2(4|4)]$$

where

$$U = q_3(1|2) + \sum_{j=4}^{\infty} q_j(1|2) \sum_{i=3}^{j-1} q_i(2|2)$$

$$V = q_3(5|4) + \sum_{j=4}^{\infty} q_j(5|4) \sum_{i=3}^{j-1} q_i(4|4) .$$

In the spirit of our previous discussion we let  $q_n(1|2) = \alpha^n q_1(1|2)$ ,  $q_n(5|4) = \beta^n q_1(5|4)$ ,  $0 < \alpha < 1$ ,  $0 < \beta < 1$ , with  $q_1(3|2) = q_2(3|2)$  and  $q_1(3|4) = q_2(3|4)$  and  $1/4 \leq q_1(2|3) = q_2(2|3) \leq 3/4$ . Our goal is to fit the estimates suggested by Table 3, namely,  $\hat{Q}(C|2) = .36$ ,  $\hat{Q}(A|2) = .05$ ,  $\hat{Q}(C|4) = .02$ ,  $\hat{Q}(A|4) = .91$ . After some numerical experimentation, the next fit was observed to be in the vicinity of the following parametric values:  $q_1(1|2) = .80$ ,  $\alpha = .75$ ,  $q_1(5|4) = .75$ ,  $\beta = .80$  with  $q_1(3|2) = .15$ ,  $q_1(3|4) = .15$  and  $q_1(2|3) = 1/3$ . For



these plausible values we obtain  $Q(C|2) = .861$ ,  $Q(A|2) = .044$ ,  $Q(C|4) = .021$ ,  $Q(A|4) = .891$  indicating a surprisingly good and rather satisfactory fit to the data.

In concluding this section it is necessary to state again that our effort for the multi-ballot model has been to develop credible exploratory models to describe two aspects of jury behavior - the overall decision-making process given the initial ballot and the ballot-to-ballot transitions that occur along the way to making jury decisions. Additional analysis is required along these lines but the lack of data makes this a somewhat esoteric exercise.

#### BIBLIOGRAPHY

- Baldwin, J. and McConville, M. (1979). Jury Trials. Oxford: Clarendon Press.
- Condorcet, N. (1785). Essai Sur La Probabilité. New York: Reprinted by Chelsea Press (1972).
- Davis, J. H. (1973). Group decision and social interaction: A theory of social decision schemes. Psychol. Review 80, 97-125.
- Davis, J. H. (1978). Group decision and procedural justice. In M. Fishbein (Ed.), Progress in Social Psychology. Erlbaum, Hillsdale, New Jersey.
- Davis, J. H., Bray, R. M. and Holt, R. (1977). The empirical study of social decision processes in juries. In J. Tapp and F. Levine (Eds.), Law, Justice, and the Individual in Society: Psychological and Legal Perspectives. New York: Holt, Rinehart & Winston.
- Davis, J. H., Kerr, N. L., Atkin, R. S., Holt, R. and Meek, D. (1975). The decision process of 6- and 12-person mock juries assigned unanimous and two-thirds majority rules. J. Personality Soc. Psychol. 32, 1-14.
- Gelfand, A. E. and Solomon, H. (1974). Modeling jury verdicts in the American legal system. J. Amer. Statist. Assoc. 69, 32-37.
- Gelfand, A. E. and Solomon, H. (1974). A group decision making model: The American jury. Tech. Report 213, Dept. of Statist., Stanford University, Stanford, California.
- Gelfand, A. E. and Solomon, H. (1975). Analyzing the decision-making process of the American jury. J. Amer. Statist. Assoc. 70, 305-309.

- Gelfand, A. E. and Solomon, H. (1977). An argument in favor of 12 member juries. Jurimetrics Jour. 17, 292-313.
- Grofman, B. (1981). Mathematical models of juror and jury decision-making: The state of the art. The Trial Process, B. D. Sales (Ed.), 305-351. New York & London: Plenum Press.
- Kalven, H. and Zeisel, H. (1966). The American Jury. Boston: Little, Brown.
- Kaye, D. (1980). Mathematical models and legal realities; reflections on the Poisson model of jury behavior. Connecticut Law Review 13, 1-15.
- Laplace, P. S. (1820). Oeuvres Completes, Vol. 7, pp. 520-530. Paris: Gauthiers-Villars, (1886).
- Pearson, K. (1978). The History of Statistics in the 17th and 18th Centuries, edited by E. S. Pearson. New York: Macmillan Publishing Company.
- Penrod, S. and Hastie, R. (1979). Models of jury decision making: A critical review. Psychol. Bull. 86, 462-492.
- Poisson, S. D. (1837). Recherches sur la probabilité des jugements en matière criminelle et en matière civile, précédées des règles générales du calcul des probabilités. Bachelier, Imprimeur-Librairie, Paris.
- Saks, M. J. (1981). Small-group decision making and complex information tasks. Federal Judicial Center, Washington, D.C.
- Walbert, T. D. (1971). Note: Effect of jury size on probability of conviction - An evaluation of Williams v. Florida. Case Western Reserve Law Review 22, 529-555.

**Supreme Court Decisions:**

- Williams v. Florida: 399US 78(1970)
- Johnson v. Louisiana: 406US 356(1972)
- Apodaca v. Oregon: 406US 404(1972)
- Ballew v. Georgia: 435US 223(1978)
- Burch v. Louisiana: 441US 130(1979)

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 335	2. GOVT ACCESSION NO. AD-A195177	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle)  JURY SIZE AND JURY VERDICTS		5. TYPE OF REPORT & PERIOD COVERED TECHNICAL REPORT
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s)  HERBERT SOLOMON		8. CONTRACT OR GRANT NUMBER(s)  N00014-76-C-0475
9. PERFORMING ORGANIZATION NAME AND ADDRESS DEPT. OF STATISTICS STANFORD UNIVERSITY STANFORD, CALIF.		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS  NR-042-267
11. CONTROLLING OFFICE NAME AND ADDRESS STATISTICS & PROBABILITY PROGRAM (code 411(SP)) OFFICE OF NAVAL RESEARCH ARLINGTON, VA 22217		12. REPORT DATE JUNE 28, 1983
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		13. NUMBER OF PAGES 38
		15. SECURITY CLASS. (of this report) UNCLASSIFIED
16. DISTRIBUTION STATEMENT (of this Report)		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)  APPROVED FOR PUBLIC RELEASE: DISTRIBUTION UNLIMITED		
18. SUPPLEMENTARY NOTES		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number)  JURY SIZE: POISSON JURY MODEL: AMERICAN JURY MODELS.		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number)  PLEASE SEE REVERSE SIDE		

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT NO. 335

ABSTRACT

The analysis of jury size and jury verdicts in criminal matters now has a long, though interrupted, history. Work on this subject in the 18th and 19th centuries by Condorcet and Laplace is discussed and the Poisson model of the 1830's is highlighted. The latter is modified to analyze the American jury experience of the 20th century. Recent U. S. Supreme Court decisions in the 1970's on jury size and jury decision-making have created a resurgence of interest especially on a comparison of six member and twelve member juries. Some comparisons of size in terms of probabilities of errors in verdicts are presented.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

END

FILMED

8-83

DTIC